A. SYSTEM DESCRIPTION

Authority: Office of Management Budget (OMB) Memorandum (M) 03–22, OMB Guidance for Implementing the Privacy Provisions of the E–Government Act of 2002 & PVR #10–Privacy Accountability and #21–Privacy Risk Management

Date of Submission: May 29, 2012 PIA ID Number: **207**

1. What type of system is this? New

1a. Is this a Federal Information Security Management Act (FISMA) reportable system? Yes

2. Full System Name, Acronym, and Release/Milestone (if appropriate):

Big Data Analytics, BDA

2a. Has the name of the system changed? No

If yes, please state the previous system name, acronym, and release/milestone (if appropriate):

3. Identify how many individuals the system contains information on

Number of Employees: More than 100,000

Number of Contractors: Over 10,000

Members of the Public: Over 1,000,000

4. Responsible Parties:

N/A

5. General Business Purpose of System

Big Data Analytics (BDA) is considered an appliance that will provide the IRS the ability to conduct advanced analytics, low latency data processing, as well as in–depth analysis of data. The BDA will be able to handle datasets and process analytics in a fast environment. Currently, the application that handles data analysis on behalf of the IRS is the Enterprise Data Access Strategy Integrated Production Model (EDAS IPM). The goal of BDA is either to augment the data analysis of EDAS IPM, or replace the application entirely. BDA will serve to perform advanced data analysis that can facilitate IRS audit selections, analyzing taxpayer filings, and more. The technology enabling this is the Massively Parallel Processing (MPP) architecture that has been designed for Business Intelligence (BI) and analytical processing. In this architecture, data is automatically partitioned across multiple 'segment' servers, and each 'segment' owns and manages a distinct portion of the overall data. The BDA is primarily made up of infrastructure and is considered a General Support System (GSS). The BDA is currently in development and is expected to be in production in September, 2012. At the time of production, it is estimated that EDAS IPM will be the only application interconnected with BDA. The BDA is made up of Greenplum hardware that was purchased from the EMC corporation. Specifically, Greenplum is a division of EMC and the product is a Commercial of the Shelf (COTS) product with modifications to conform with IRS IRM requirements.

6. Has a PIA for this system, application, or database been submitted previously to the Office of Privacy Compliance? (*If you do not know, please contact* *Privacy *and request a search*) Yes

6a. If Yes, please indicate the date the latest PIA was approved: 12/12/2011

6b. If Yes, please indicate which of the following changes occurred to require this update.

● System Change (1 or more of the 9 examples listed in OMB 03–22 applies)
(refer to PIA Training Reference Guide for the list of system changes) No

● System is undergoing Security Assessment and Authorization Yes

6c. State any changes that have occurred to the system since the last PIA

Milestone 2/3/4a of the ELC.

7. If this system has an Exhibit 53 or Exhibit 300 please provide the Unique Project Identifier (UPI) number (XXX–XX–XX–XX–XX–XXXX–XX). Otherwise, enter the word 'none' or 'NA'. NA

B. DATA CATEGORIZATION

Authority: OMB M 03–22 & PVR #23–PII Management

8. Does this system collect, display, store, maintain or disseminate Personally Identifiable Information (PII)? Yes

8a. If No, what types of information does the system collect, display, store, maintain or disseminate?

9. Indicate the category that best describes the source that provides or originates the PII collected, displayed, stored, maintained or disseminated by this system. Most common categories follow:

Taxpayers/Public/Tax Systems	Yes	
Employees/Personnel/HR Systems	Yes	
		Other Source:
Other	Yes	Contractors

10. Indicate all of the types of PII collected, displayed, stored, maintained or disseminated by this system. Then state if the PII collected is on the Public and/or Employees. Most common fields follow:

TYPE OF PII	Collected?	On Public?	On IRS Employees or Contractors?
Name	Yes	Yes	Yes
Social Security Number (SSN)	Yes	Yes	Yes
Tax Payer ID Number (TIN)	Yes	Yes	Yes
Address	Yes	Yes	Yes
Date of Birth	Yes	Yes	Yes

Additional Types of PII: No

No Other PII Records found.

10a. Briefly describe the PII available in the system referred to in question 10 above.

The PII data available is contained on the various tax forms that are filed with the IRS by members of the general public.

If you answered **Yes** to Social Security Number (SSN) in question 10, answer **10b**, **10c**, and **10d**.

10b. Cite the authority that allows this system to contain SSN's? (e.g. specific regulations, statutes, etc.)

26 USC 3402, 3406, 1441 and IRC 6109

10c. What alternative solution to the use of the SSN has/or will be applied to this system? (e.g. masking, truncation, alternative identifier)

There is no alternative to the use of the SSN. The SSN is the significant part of the data being processed.

10d. Describe the planned mitigation strategy and forecasted implementation date to mitigate or eliminate the use of Social Security Numbers on this system?

There is no plan to eliminate the use of the SSN on the system.

11. Describe in detail the system's audit trail. State what data elements and fields are collected. Include employee log–in information. If the system does not have audit capabilities, explain why an audit trail is not needed.

The Design and Development phase has just begun. BDA will engage the ESAT office to develop the BDA ESAT Audit Plan. Audit records will capture all of the required elements contained with the IRS IRMs such as account logon, activity of admin users, as well as failed logon attempts.

11a. Does the audit trail contain the audit trail elements as required in current IRM 10.8.3 *Audit Logging Security Standards*? Yes

12. What are the sources of the PII in the system? Please indicate specific sources:

a. IRS files and databases: Yes

If **Yes**, the system(s) are listed below:

System Name	Current PIA?	PIA Approval Date	SA & A?	Authorization Date
EDAS IPM	Yes	02/21/2012	Yes	08/01/2011
EIP	Yes	03/28/2011	Yes	09/13/2011

b. Other federal agency or agencies: No
If Yes, please list the agency (or agencies) below:

c. State and local agency or agencies: No
If Yes, please list the agency (or agencies) below:

d. Third party sources: No
If yes, the third party sources that were used are:

e. Taxpayers (such as the 1040): Yes

f. Employees (such as the I–9): No

g. Other: No **If Yes*, specify*:**

C. PURPOSE OF COLLECTION

Authorities: OMB M 03–22 & Internal Revenue Manual (IRM) 10.8.8, IT Security, Live Data Protection Policy & PVR #16, Acceptable Use

13. What is the business need for the collection of PII in this system? Be specific.

The BDA project is an infrastructure project that will provide the IRS with a massively parallel processing capability to support many projects that have a need for case identification, selection, prioritization and delivery and compliance and decision analytics. BDA will serve to perform advanced data analysis that can facilitate IRS audit selections, analyzing taxpayer filings, and more.

D. PII USAGE

Authority: OMB M 03–22 & PVR #16, Acceptable Use

14. What is the specific use(s) of the PII?

To conduct tax administration	Yes	
To provide taxpayer services	No	
To collect demographic data	No	
For employee purposes	No	
		If other, what is the use?
Other:	No	

E. INFORMATION DISSEMINATION

Authority: OMB M 03–22 & PVR #14–Privacy Notice and #19–Authorizations

15. Will the information be shared outside the IRS? (for purposes such as computer matching, statistical purposes, etc.) No

15a. If yes, with whom will the information be shared? The specific parties are listed below:

	Yes/No	Who?	ISA OR MOU**?
Other federal agency (–ies)			
State and local agency (–ies)			
Third party sources			
Other:			

** Inter–agency agreement (ISA) or Memorandum of Understanding (MOU)

16. Does this system host a website for purposes of interacting with the public? No

17. Does the website use any means to track visitors' activity on the Internet?

If yes, please indicate means:

	YES/NO	AUTHORITY
Persistent Cookies		
Web Beacons		
Session Cookies		

If other, specify:

Other: ____________ ____________

F. INDIVIDUAL CONSENT

Authority: OMB M 03–22 & PVR #15–Consent and #18–Individual Rights

18. Do individuals have the opportunity to decline to provide information or to consent to particular uses of the information? Not Applicable

18a. If Yes, how is their permission granted?

19. Does the system ensure "due process" by allowing affected parties to respond to any negative determination, prior to final action? Not Applicable

19a. If Yes, how does the system ensure "due process"?

20. Did any of the PII provided to this system originate from any IRS issued forms? Yes

20a. If Yes, please provide the corresponding form(s) number and name of the form.

Form Number	Form Name
1040	U.S. Individual Income Tax Return
1120	U.S. Corporation Income Tax Return
941	Employer's Quarterly Federal Tax Return
940	Employer's Annual Federal Unemployment Tax Return
1065	U.S. Return of Partnership Income
1099	Information Return
W2	Federal Income Tax Withheld
Many more forms will be supplied upon request	Too numerous to list

20b. If No, how was consent granted?

Written consent ____________

Website Opt In or Out option ____________

Published System of Records Notice in the Federal Register ____________

Other: ____________

G. INFORMATION PROTECTIONS

Authority: OMB M 03–22 & PVR #9–Privacy as Part of the Development Life Cycle, #11–Privacy Assurance, #12–Privacy Education and Training, #17–PII Data Quality, #20–Safeguards and #22–Security Measures

21. Identify the owner and operator of the system: IRS Owned and Operated

21a. If Contractor operated, has the business unit provided appropriate notification to execute the annual security review of the contractors, when required?

22. The following people have use of the system with the level of access specified:

	Yes/No	Access Level
IRS Employees:	Yes	
Users		Read Only
Managers		Read Only
System Administrators		Read Only
Developers		Read Write
Contractors:	No	
Contractor Users		
Contractor System Administrators		
Contractor Developers		
Other:	No	

If you answered yes to contractors, please answer 22a. ***(All contractor/contractor employees must hold at minimum, a "Moderate Risk" Background Investigation if they have access to IRS owned SBU/PII data.)***

22a. If the contractors or contractor employees act as System Administrators or have "Root Access", does that person hold a properly adjudicated "High Level"background investigation?

23. How is access to the PII determined and by whom?

Access to BDA is granted to client applications that have a business need to connect to BDA. There are no individual users. Access is determined by the Data Strategy Advisory Board (DSAB). The perspective client applications must present a current IATO/ATO, and a Change Request (CR) that details the business need and the connection information.

24. How will each data element of SBU/PII be verified for accuracy, timeliness, and completeness?

The BDA system receives data from the EDAS–IPM system which has its own verification process for data accuracy, timeliness, completeness and therefore BDA assumes that the data is accurate, timely, and complete when it is provided by EDAS–IPM.

25. Are these records covered under the General Records Schedule (GRS), or have a National Archives and Records Administration (NARA) archivist approved a Record Control Schedule (RCS) for the retention and destruction of official agency records stored in this system? No

25a. If Yes, how long are the records required to be held under the corresponding RCS and how are they disposed of? In your response, please include the complete IRM number 1.15.XX and specific item number and title.

If No, how long are you proposing to retain the records? Please note, if you answered no, you must contact the IRS Records and Information Management Program to initiate records retention scheduling before you dispose of any records in this system.

BDA is non–recordkeeping and does not require a National Archives–approved records control schedule to affect data disposition. BDA is an infrastructure project that will provide low latency data processing of recordkeeping data appropriately maintained and scheduled in the context of those source systems. BDA's design and development phase has just begun and the PIA will be updated as necessary.

26. Describe how the PII data in this system is secured, including appropriate administrative and technical controls utilized.

Access to BDA is limited to those with appropriate need and authority. Also, the BDA has various technical controls such as password complexity when logging on to the application. There is also a warning banner, account lockout (after a specified number of failed logons), and more. BDA also provides for separation of duties. BDA provides for the separation of duties through role based privileges that separate sensitive responsibilities. User roles are separated in order to limit conflicts of interest in the responsibilities and interests of individuals, therefore ensuring a single user does not have privileges to perform multiple conflicting security functions. Separation of duties is enforced through roles that are assigned to each user. General users cannot access functions available to Managers or SAs. Users are assigned access authorizations by their manager or manager proxy. The system administrator would review the assigned authorization and grant/deny privileges associated with the authorization. A separation of duty also exists between the development and production environments. In the development environment, the application developers are responsible for making the required changes to application code, but they are not authorized to move application code into the production environment. The SAs are responsible for migrating the code into production, and development staff is restricted from being able to perform this function. Lastly, a manager cannot approve their own requests.

26a. Next, explain how the data is protected in the system at rest, in flight, or in transition.

Access to BDA is limited to those with appropriate need and authority. Also, the BDA has various technical controls such as password complexity when logging on to the application. There is also a warning banner, account lockout (after a specified number of failed logons), and more. BDA also provides for separation of duties. BDA provides for the separation of duties through role based privileges that separate sensitive responsibilities. User roles are separated in order to limit conflicts of interest in the responsibilities and interests of individuals, therefore ensuring a single user does not have privileges to perform multiple conflicting security functions. Separation of duties is enforced through roles that are assigned to each user. General users cannot access functions available to Managers or SAs. Users are assigned access authorizations by their manager or manager proxy. The system administrator would review the assigned authorization and grant/deny privileges associated with the authorization. A separation of duty also exists between the development and production environments. In the development environment, the application developers are responsible for making the required changes to application code, but they are not authorized to move application code into the production environment. The SAs are responsible for migrating the code into production, and development staff is restricted from being able to perform this function. Lastly, a manager cannot approve their own requests.

27. Has a risk assessment (e.g., SA&A) been conducted on the system to ensure that appropriate security controls have been identified and implemented to protect against known risks to the confidentiality, integrity and availability of the PII? No

28. Describe the monitoring/evaluating activities undertaken on a regular basis to ensure that controls continue to work properly in safeguarding the PII.

Continuous Monitoring (eCM) is performed annually to determine if selected System Security Plan (SSP) controls are operating as intended. The Security Assessment and Authorization (SA&A) process is conducted on a three year cycle whereby all application information and control descriptions are updated and tested to ensure that the controls continue to work properly in safeguarding the PII. Findings from the SA&A are detailed in the Security Assessment Report (SAR) leading to the mitigation of the findings.

29. Is testing performed, in accordance with Internal Revenue Manual (IRM) 10.8.8 – *IT Security, Live Data Protection Policy*? Not Applicable

29a. Has approval been received from the Office of Privacy Compliance to use Live Data in testing (*if appropriate)*?

29b. If you have received permission from the Office of Privacy Compliance to use Live Data, when was the approval granted?

H. PRIVACY ACT & SYSTEM OF RECORDS

Under the statute, any employee who knowingly and willfully maintains a system of records without meeting the Privacy Act notice requirements is guilty of a misdemeanor and may be fined up to $5000.

Authority: OMB M 03–22 & Privacy Act, 5 U.S.C. 552a (e) (4) & PVR #13–Transparency

30. Are 10 or more records containing PII maintained/stored/transmitted through this system? Yes

31. Are records on the system retrieved by any identifier for an individual? (Examples of identifiers include but are not limited to Name, SSN, Photograph, IP Address) Yes

31a. If YES, the System of Records Notice(s) (SORN) published in the Federal Register adequately describes the records as required by the Privacy Act? Enter the SORN number and the complete name of the SORN.

SORNS Number	SORNS Name
Treasury/IRS 24.030	Individual Master File
Treasury/IRS 42.021	Compliance Programs and Project Files
Treasury/IRS 24.046	Business Master File
Treasury/IRS 34.037	IRS Audit Trail and Security Records System
Treasury/IRS 25.046	BMF

Comments

I. ANALYSIS

Authority: OMB M 03–22 & PVR #21– Privacy Risk Management

32. What choices were made or actions taken regarding this IT system or collection of information as a result of preparing the PIA?

Resulted in the removal of PII from the system (e.g., SSN use reduced/eliminated)	No
Provided viable alternatives to the use of PII within the system	No
New privacy measures have been considered/implemented	No
Other:	No

32a. If Yes to any of the above, please describe:

N/A

View other PIAs on IRS.gov